Confidence in the Press and Political Polarization

Ari Gelbard

Table of contents

This document aims to answer the question how has trust in the press changed over time? And has this increased with increased political polarization?

Order:

- Data Exploration

  - Trust in the media
  - Political Ideology and Polarization
  - Press Confidence and Political Ideology/Polarization

- Hypothesis Testing

Loading Packages and Data

```r
### Load packages
library(tidyverse)
library(readxl)
library(zoo)
library(boot)

### Load data
gss_raw <- read_excel("data/gss_press.xlsx")
```

Cleaning Data

```r
gss <- gss_raw %>%
  mutate(partyid  = ifelse(partyid %in% c(".n:  No answer",".d:  Do not Know/Cannot Choose")
                           NA, partyid),
         conpress = ifelse(conpress %in% c("ONLY SOME","HARDLY ANY","A GREAT DEAL"),
                           conpress,NA),
         polviews = ifelse(polviews %in% c(".i:  Inapplicable",".d:  Do not Know/Cannot Choos
                                           ".n:  No answer",".s:  Skipped on Web"),
                           NA, polviews)) %>%
  mutate(party3 = case_when(
    partyid %in% c("Strong democrat",
                   "Not very strong democrat") ~ "Democrat",
    partyid %in% c("Strong republican",
                   "Not very strong republican") ~ "Republican",
    partyid %in% c("Independent (neither, no response)",
                   "Independent, close to republican",
                   "Independent, close to democrat") ~ "Independent")) %>%
  mutate(year = as.numeric(year),
         id_  = as.numeric(id_))

gss_balanced <- gss %>% drop_na()
```

Data Analysis

In this section, I want to explore the GSS variables we have to see if we can observe any interesting trends.

Press Confidence

Our confidence in the press variables is conpress, which asked the respondent whether have a great deal of confidence, only some confidence, or hardly any confidence in the press? Lets plot it over time to see how opinions have evolved.

```r
# Get counts and proportion for each answer per years
conpress <- gss %>%
  group_by(year,conpress) %>%
  summarise(n = n(), .groups="drop") %>%
  group_by(year) %>%
  drop_na() %>%
  mutate(prop = n / sum(n))

ggplot(conpress, aes(x = year, y = prop, color = conpress)) +
  geom_line(linewidth = 1.2) +
  scale_y_continuous(labels = scales::percent) +
  labs(
    title = "Does the respondent have a great deal of confidence,\nonly some confidence,or h
    x = "Year",
    y = "Proportion of Respondents",
    color = "Response"
  ) +
  theme_bw()
```

Does the respondent have a great deal of confidence,
only some confidence,or hardly any confidence in the press?
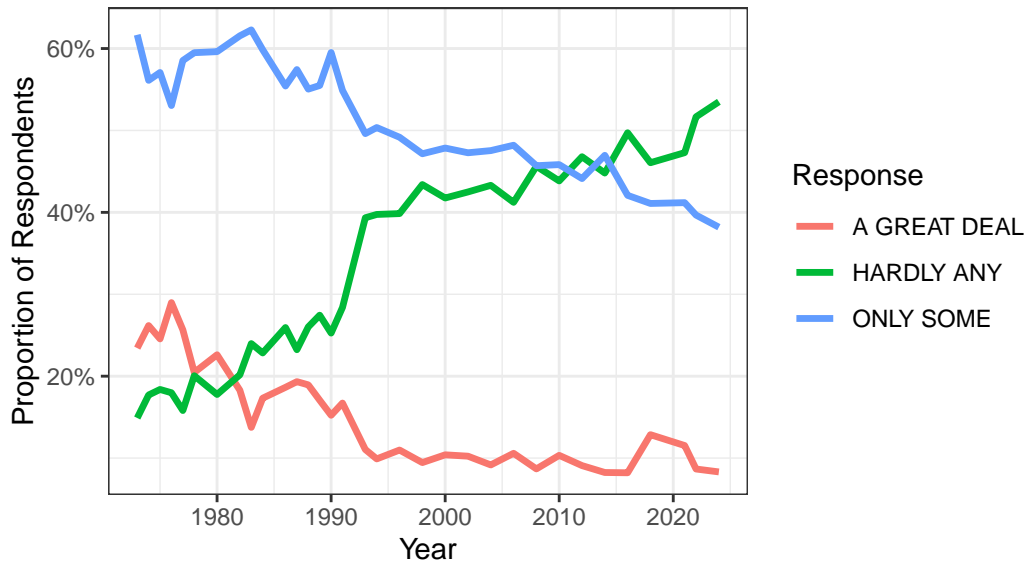
It seems pretty evident that confidence in the press has decreased significantly over time. In fact, the majority of respondents in 2024 have hardly any confidence in the press.

Political Polarization

Now lets look at political polarization. The two variables we have are as follows:

- partyid: Generally speaking, does the respondent think of themselves as a Republican, Democrat, Independent, or what?
- polviews: On a 7 point scale between extremely liberal and extremely conservative, where would you place yourself on this scale?

Lets start by looking at how the number of respondents for these variables has evolved over time by category. For party identification, we combine the results into 3 categories: Democrat, Independent, and Republican.

```
gss_party_agg <- gss %>%
  group_by(year, party3) %>%
  summarise(n = n(), .groups="drop") %>%
  group_by(year) %>%
  drop_na() %>%
  mutate(prop = n / sum(n))
```

```
ggplot(gss_party_agg, aes(x = year, y = prop, color = party3)) +
  geom_line(linewidth = 1.2) +
  scale_y_continuous(labels = scales::percent) +
  scale_color_manual(name = "Party ID",
                     values = c("Democrat" = "steelblue2",
                                "Independent" = "gray40",
                                "Republican" = "firebrick3")) +
  labs(
    title = "Party Identification Over Time",
    x = "Year",
    y = "Proportion of Respondents",
    color = "Party ID"
  ) +
  theme_bw()
```
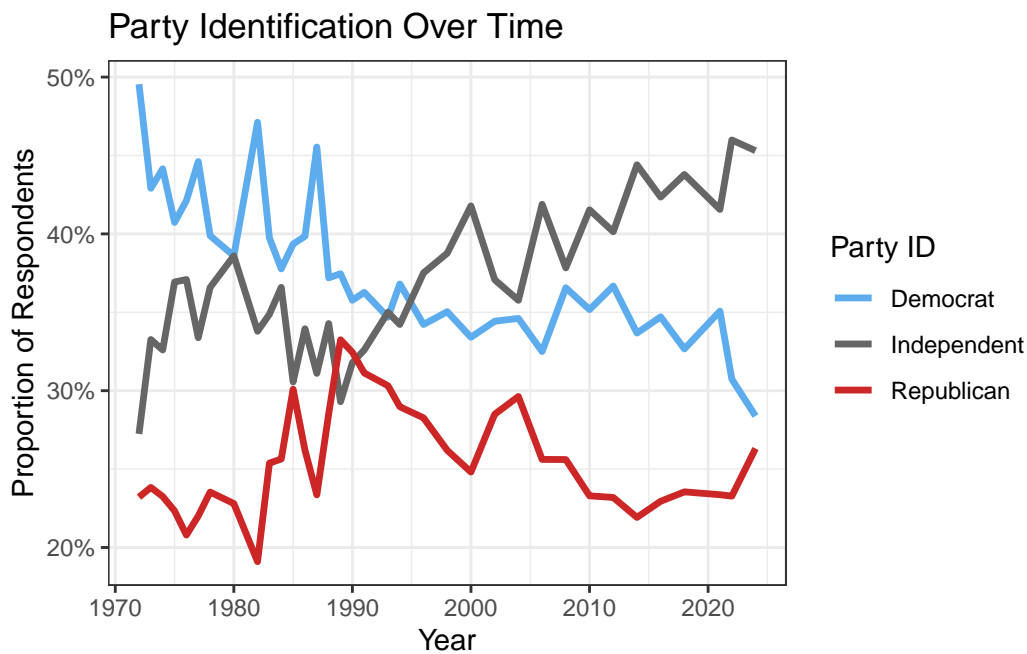

It appears that over time, the number of people who identify as republican has increased slightly, the number of people who identify as democrat has decreased, and the number of people who identify as independent has increased significantly. This could indicate a dissatisfaction with the current political parties, leading more people to identify as independent. However, this doesn't give us much insight into polarization.

Let's look at political views over time since this corresponds more closely to people's beliefs than simple party identification.

```r
polviews <- gss %>%
  mutate(polviews = factor(polviews,
                           levels = c("Extremely liberal",
                                      "Liberal",
                                      "Slightly liberal",
                                      "Moderate, middle of the road",
                                      "Slightly conservative",
                                      "Conservative",
                                      "Extremely conservative"))) %>%
  group_by(year,polviews) %>%
  summarise(n = n(), .groups="drop") %>%
  group_by(year) %>%
  drop_na() %>%
  mutate(prop = n / sum(n))

ggplot(polviews, aes(x = year, y = prop, color = polviews)) +
  geom_line(linewidth = 1.2) +
  scale_y_continuous(labels = scales::percent) +
  scale_color_manual(name = "Political Ideology",
                     values = c("Extremely liberal" = "blue3",
                                "Liberal" = "royalblue",
                                "Slightly liberal" = "lightskyblue",
                                "Moderate, middle of the road" = "gray40",
                                "Slightly conservative" = "pink",
                                "Conservative" = "tomato",
                                "Extremely conservative" = "firebrick3")) +
  labs(
    title = "Political Ideology of Respondents Over Time",
    x = "Year",
    y = "Proportion of Respondents",
    color = "Political Ideology"
  ) +
  theme_bw()
```
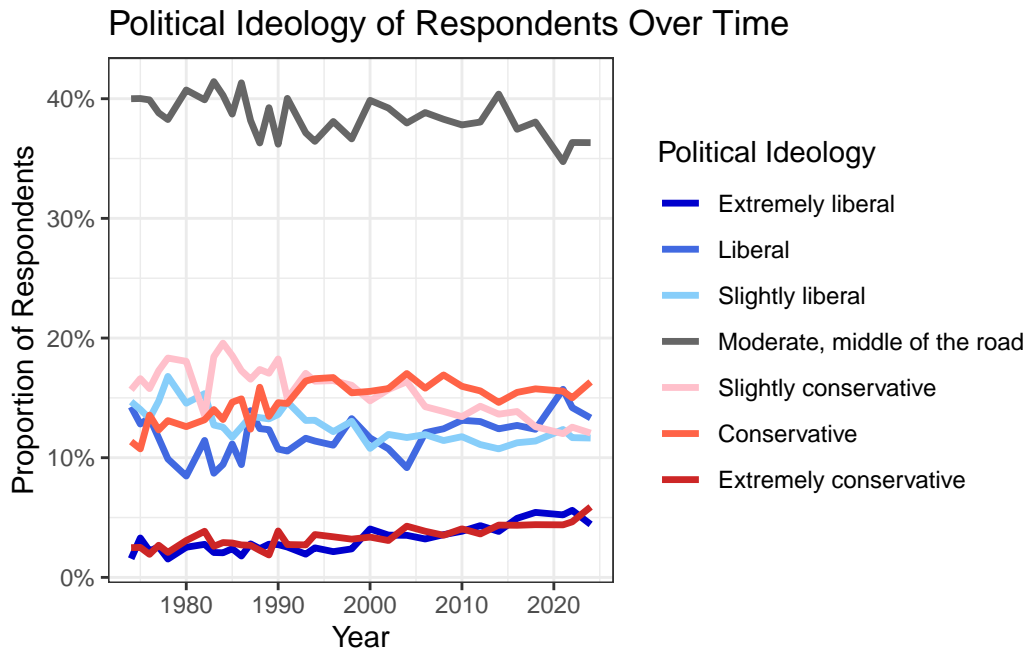
Political Ideology of Respondents Over Time

Most categories appear to have remained fairly stagnant. However there does seem to be a slight increase in the proportion of people who consider themselves extremely liberal and extremely conservative from about 2% to about 5% for both categories. We can see this trend more concretely if we look at the standard deviation of responses. We can do this by converting them back into numerical variables that range from 1 to 7, and then calculating the standard deviation in each year and plotting it.

```
gss <- gss %>%
  mutate(partyid_num = ifelse(partyid == "Strong democrat", 1,
                       ifelse(partyid == "Not very strong democrat", 2,
                       ifelse(partyid == "Independent, close to democrat", 3,
                       ifelse(partyid == "Independent (neither, no resp
                       ifelse(partyid == "Independent, close to
                       ifelse(partyid == "Not very strong
                       ifelse(partyid == "Strong r

gss <- gss %>%
  mutate(polviews_num = ifelse(polviews == "Extremely liberal", 1,
                        ifelse(polviews == "Liberal", 2,
                        ifelse(polviews == "Slightly liberal", 3,
                        ifelse(polviews == "Moderate, middle of the road
                        ifelse(polviews == "Slightly conservative
                        ifelse(polviews == "Conservative"
```

```
                                                            ifelse(polviews == "Extreme

spread_by_year <- gss %>%
  group_by(year) %>%
  summarise(
    sd_pol = sd(polviews_num, na.rm = TRUE),
    sd_pid = sd(partyid_num, na.rm = TRUE),
    n = sum(!is.na(polviews_num))
  )

ggplot(data=spread_by_year,aes(x=year)) +
  geom_line(linewidth = 1.2, aes(y=sd_pol, color="Political Views")) +
  geom_line(linewidth = 1.2, aes(y=sd_pid, color="Party ID")) +
  scale_color_manual(name = "Variable",
                     values = c("Political Views" = "firebrick",
                                "Party ID" = "deepskyblue4")) +
  labs(title = "Standard Deviation of Political Views Over Time",
       x = "Year",
       y = "Standard Deviation") +
  theme_minimal()
```
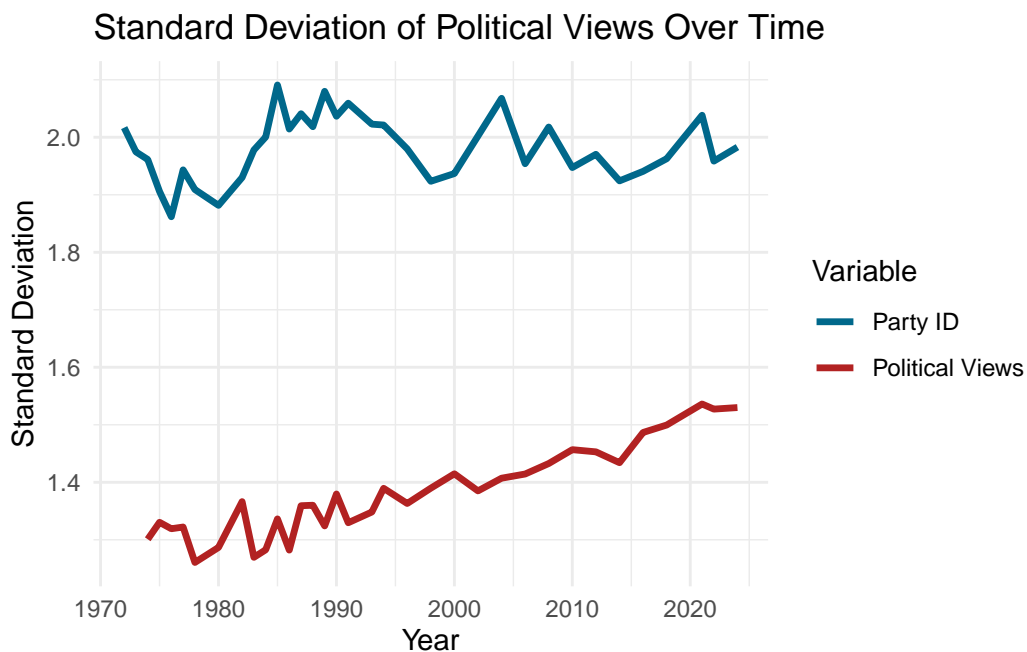

We can see that the standard deviation for Party ID hasn't changed much, but the standard deviation of political views has increased from about 1.3 to over 1.5. But is this a significant

change?

To test whether the standard deviation of political ideology has actually increased over time, we can do a bootstrapping test of the standard deviation trend. We can do this by estimating a bootstrapping standard deviation for the periods before and after 1998, and then we can take the difference of the resulting 95% confidence intervals to get an interval for the change.

```r
N <- 5000

# function to compute SD
boot_sd <- function(data, idx) {
  sd(data[idx], na.rm = TRUE)
}

early <- gss$polviews_num[gss$year < 1998]
late  <- gss$polviews_num[gss$year >= 1998]

boot_early <- boot(early, boot_sd, R = N)
boot_late  <- boot(late, boot_sd, R = N)

boot.ci(boot_late, type="perc")
```

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 5000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_late, type = "perc")

Intervals :
Level     Percentile
95%   ( 1.455,  1.474 )
Calculations and Intervals on Original Scale
```

```r
boot.ci(boot_early, type="perc")
```

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 5000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_early, type = "perc")

Intervals :
```

```
Level     Percentile
95%   ( 1.325,  1.346 )
Calculations and Intervals on Original Scale
```

```
# Compare distributions or compute CI of difference
sd_diff <- boot_late$t - boot_early$t
quantile(sd_diff, c(.025, .975))
```

```
     2.5%     97.5%
0.1150033 0.1429368
```

Since the difference in the 95% bootstrap confidence interval is positive, we can conclude with 95% certainty that the standard deviation of political views increased by between 0.116 and 0.142 standard deviations.

Based on these results, we can confidently say that political polarization has increased over time.
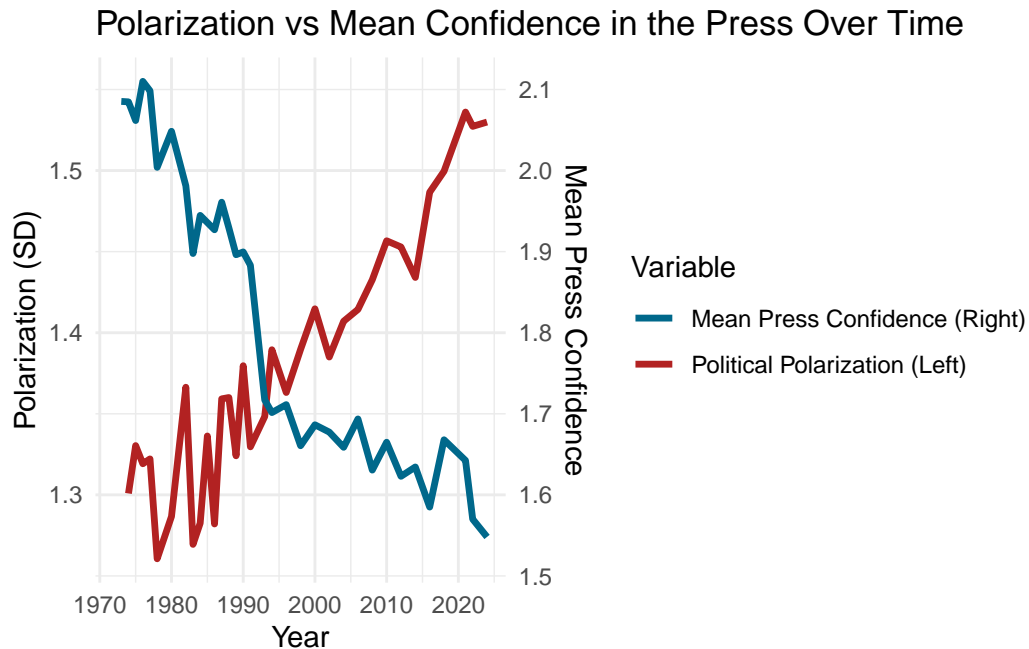
Media Confidence and Political Ideology

Now let's investigate how political polarization and media confidence relate. First, we can just plot the standard deviation of political ideology from the previous section with mean press confidence over time.

```
press_pol <- gss %>%
  mutate(conpress_num = case_when(
    conpress == "HARDLY ANY" ~ 1,
    conpress == "ONLY SOME" ~ 2,
    conpress == "A GREAT DEAL" ~ 3)) %>%
  group_by(year) %>%
  summarise(mean_conf = mean(conpress_num, na.rm = TRUE)) %>%
  right_join(spread_by_year, by = "year") %>%
  arrange(year) %>%
  mutate(mean_conf = na.approx(mean_conf, x = year, na.rm = FALSE),
         sd_pol    = na.approx(sd_pol,    x = year, na.rm = FALSE))

ggplot(press_pol, aes(x = year)) +
  geom_line(aes(y = sd_pol, color = "Political Polarization (Left)"), linewidth=1.2) +
  geom_line(aes(y = 0.5 * mean_conf + 0.5, color = "Mean Press Confidence (Right)"), linewidt
  scale_y_continuous(name = "Polarization (SD)",
                     sec.axis = sec_axis(~ (. - .5) / .5, name = "Mean Press Confidence")) +
```

```
scale_color_manual(values = c("Political Polarization (Left)" = "firebrick",
                              "Mean Press Confidence (Right)" = "deepskyblue4")) +
labs(title="Polarization vs Mean Confidence in the Press Over Time",
     x="Year",
     color="Variable") +
theme_minimal()
```

Polarization vs Mean Confidence in the Press Over Time

Based on this chart, we can see more clearly that political polarization has been increasing while mean press confidence has been decreasing. But how has average confidence in the press evolved over time for different political ideologies?

```
mean_by_pol <- gss %>%
  mutate(conpress_num = case_when(
    conpress == "HARDLY ANY" ~ 1,
    conpress == "ONLY SOME" ~ 2,
    conpress == "A GREAT DEAL" ~ 3
  )) %>%
  mutate(polviews_cat = polviews) %>%
  group_by(year, polviews_cat) %>%
  summarise(mean_conf = mean(conpress_num, na.rm=TRUE), .groups="drop") %>%
  filter(!is.na(polviews_cat)) %>%
  group_by(polviews_cat) %>%
  arrange(year) %>%
```

```
    mutate(mean_conf = na.approx(mean_conf, x = year, na.rm = FALSE)) %>%
  ungroup() %>%
  mutate(polviews_cat = factor(polviews_cat,
                           levels = c("Extremely liberal",
                                      "Liberal",
                                      "Slightly liberal",
                                      "Moderate, middle of the road",
                                      "Slightly conservative",
                                      "Conservative",
                                      "Extremely conservative")))

ggplot(mean_by_pol, aes(x = year, y = mean_conf, group = polviews_cat, color = polviews_cat))
  geom_line(linewidth = 0.9) +
  scale_color_manual(name = "Political Ideology",
                     values = c("Extremely liberal" = "navyblue",
                                "Liberal" = "dodgerblue2",
                                "Slightly liberal" = "lightskyblue2",
                                "Moderate, middle of the road" = "gray40",
                                "Slightly conservative" = "pink",
                                "Conservative" = "firebrick1",
                                "Extremely conservative" = "firebrick4")) +
  labs(
    title = "Mean Press Confidence Across Ideological Groups",
    x = "Year",
    y = "Mean Confidence",
    color = "Political Ideology"
  ) +
  theme_bw()
```
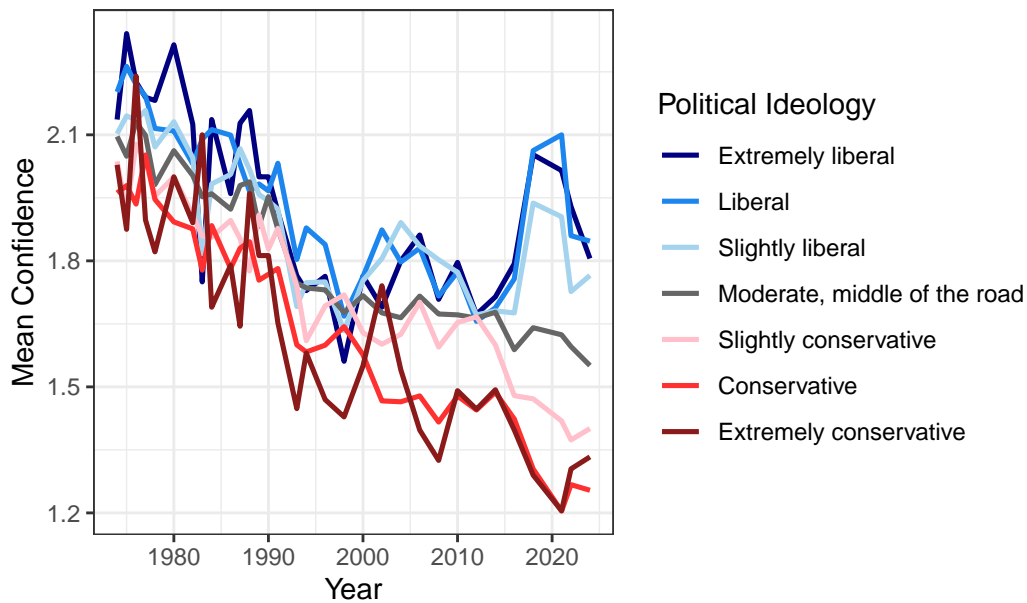
Mean Press Confidence Across Ideological Groups

All ideologies seem to have been loosing confidence in the press over time. There was a large spike in the late 2010s among liberals, but their confidence has still decreased over all and appears like it may continue to decrease.

We can test whether the distributions for political identification and confidence in the press are independent using a Chi-Squared test for independence. The null and alternative hypotheses for this test are as follows

H_0 : The distributions for political views and confidence in the press **are** independent.

H_a : The distributions for political views and confidence in the press **are not** independent.

```
chisq.test(table(gss$polviews, gss$conpress))
```

```
    Pearson's Chi-squared test

data:  table(gss$polviews, gss$conpress)
X-squared = 1551.1, df = 12, p-value < 2.2e-16
```

The resulting p-value is less than 0.05 which means that we can reject the null hypothesis at the 5% level. In other words, we have sufficient statistical evidence to conclude that the distributions of political views and confidence in the press are not independent.

We can also check if the dependence has increased over time by rerunning the Chi-Squared test for every year and plotting the results.

```r
gss_chisq <- gss %>%
  filter(!is.na(polviews), !is.na(conpress)) %>%
  mutate(polviews = factor(polviews),
         conpress = factor(conpress))

chi_by_year <- gss_chisq %>%
  group_by(year) %>%
  summarise(
    chisq = list(chisq.test(table(polviews, conpress)))
  ) %>%
  mutate(
    statistic = map_dbl(chisq, "statistic"),
    p_value   = map_dbl(chisq, "p.value"),
    df        = map_dbl(chisq, ~ .x$parameter)
  ) %>%
  select(-chisq)

ggplot(chi_by_year, aes(x = year, y = statistic)) +
  geom_line(linewidth = 1.2, color = "firebrick") +
  labs(
    title = "Chi-square Statistic Over Time",
    subtitle = "Association Between Political Ideology and Press Confidence",
    x = "Year",
    y = "Chi-square Statistic"
  ) +
  theme_minimal(base_size = 14)
```

Chi−square Statistic Over Time

Association Between Political Ideology and Press Confid
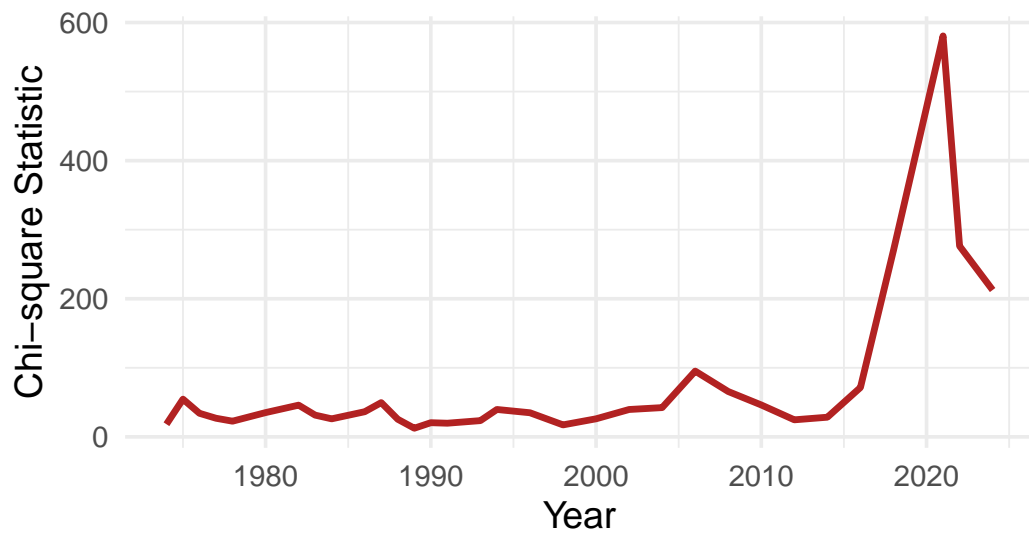

```
ggplot(chi_by_year, aes(x = year, y = p_value)) +
  geom_line(linewidth = 1.2, color = "steelblue") +
  geom_hline(yintercept = 0.05, linetype = "dashed") +
  scale_y_continuous(trans = "log10") +
  labs(
    title = "P-values for Chi-square Tests Over Time",
    subtitle = "(log-scale axis)",
    x = "Year",
    y = "P-value"
  ) +
  theme_minimal(base_size = 14)
```

P−values for Chi−square Tests Over Time
(log−scale axis)
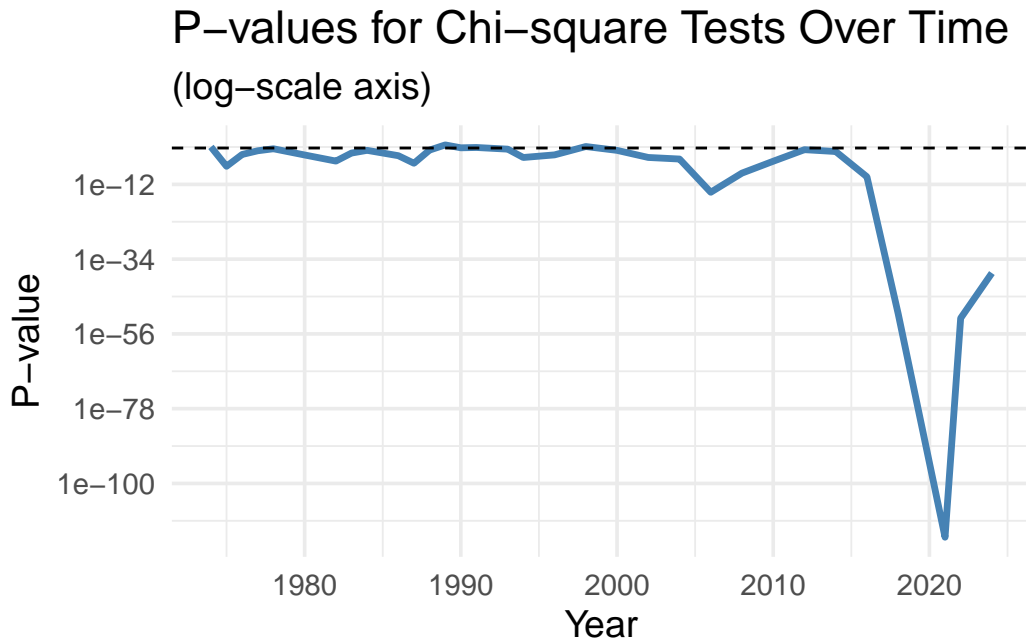
We can see from the p-value plot that most p-values are below the 5% level. This means that we can reject the null hypothesis at the 5% level for most years in the sample. Furthermore, there is a large spike after 2016, which corresponds with the beginning of President Trump's first term.

Conclusion

Although we cannot confidently say that increased political polarization is a cause of or a result of a decrease in press confidence over time, we can say with certainty that the two are closely related and have been covarying, particularly in the past few years.